

AWS re:Invent

NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

AIM214

Learn how Black Knight is using AI to accelerate mortgage workflows

Aditya Krishnan

General Manager,
AWS Machine Learning, Computer Vision AI
Amazon Web Services

Frank Poiesz

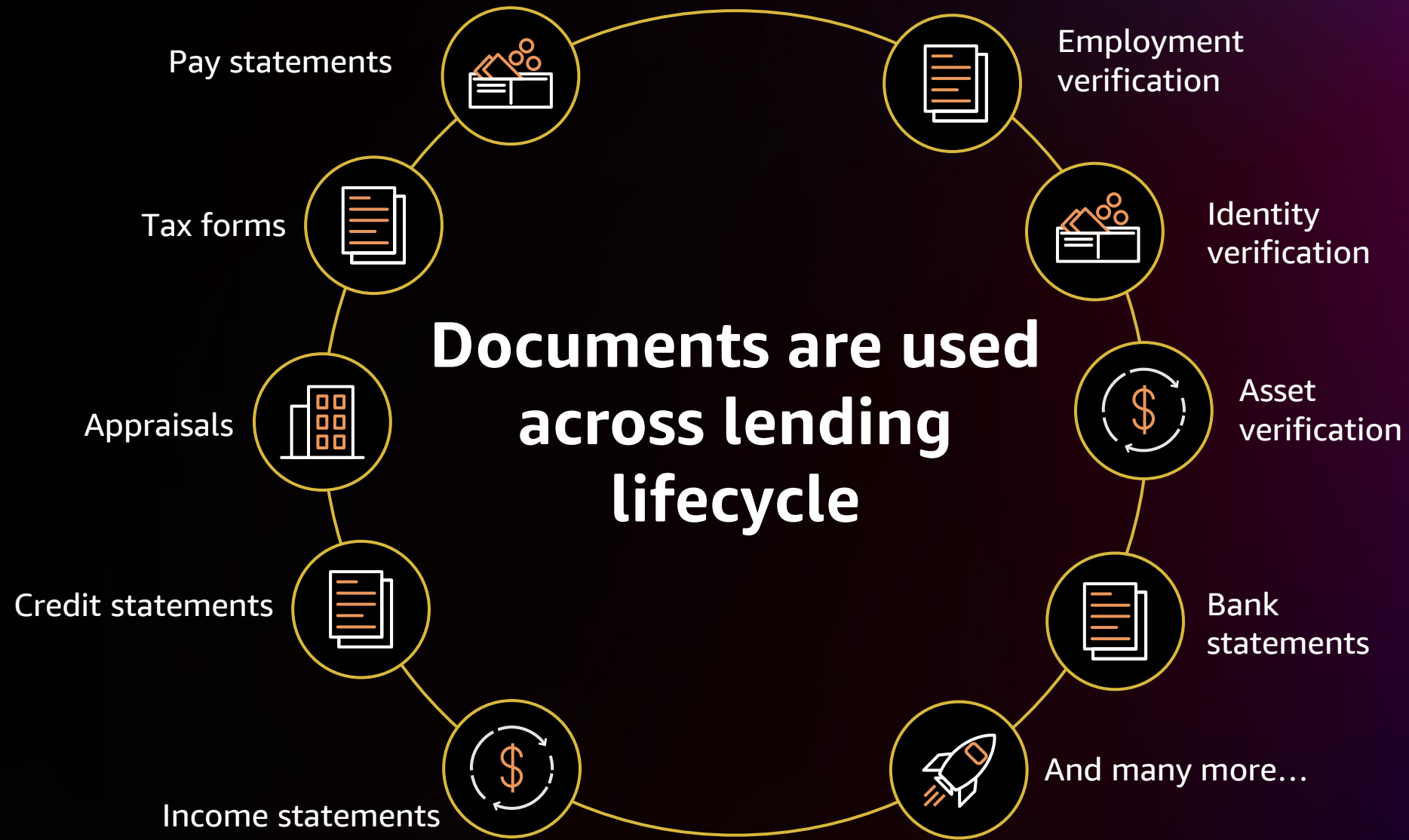
Business Strategy Director, Origination Technology
Black Knight



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

1. Introduction to intelligent document processing
2. Challenges with legacy document processing
3. Overview of Amazon Textract capability
4. Black Knight – Case study
5. Helpful links
6. Questions



Legacy document processes do not meet today's needs



Legacy OCR and manual processes are time-consuming, error-prone, and expensive



Manual processes do not scale easily with document volume



Difficult to find useful information needed for business decisions

Faster document processing shortens decision cycles;
this enables you to serve more customers and have people do higher-value tasks

The hidden costs of legacy systems

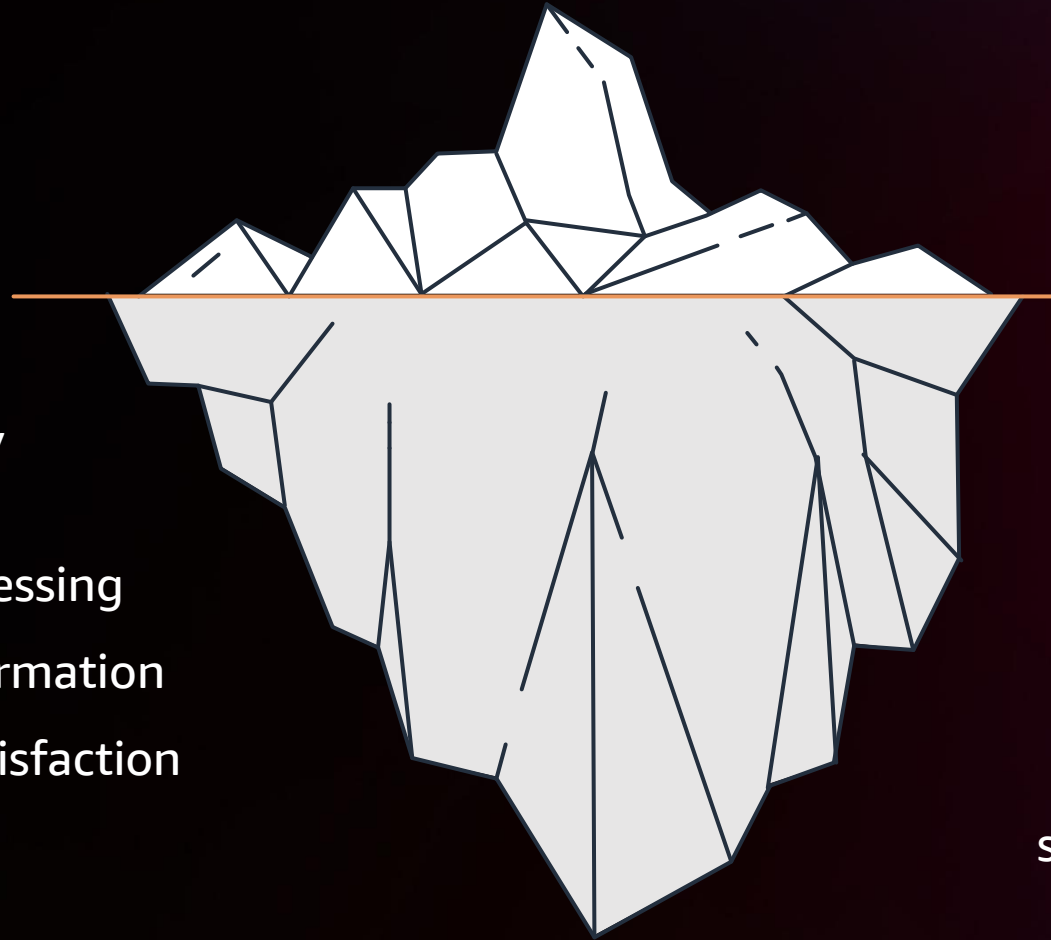
Lost revenue

Limited capacity of legacy systems blocks growth

Extremely slow data processing

Higher barriers to transformation

Errors lower customer satisfaction and drive churn



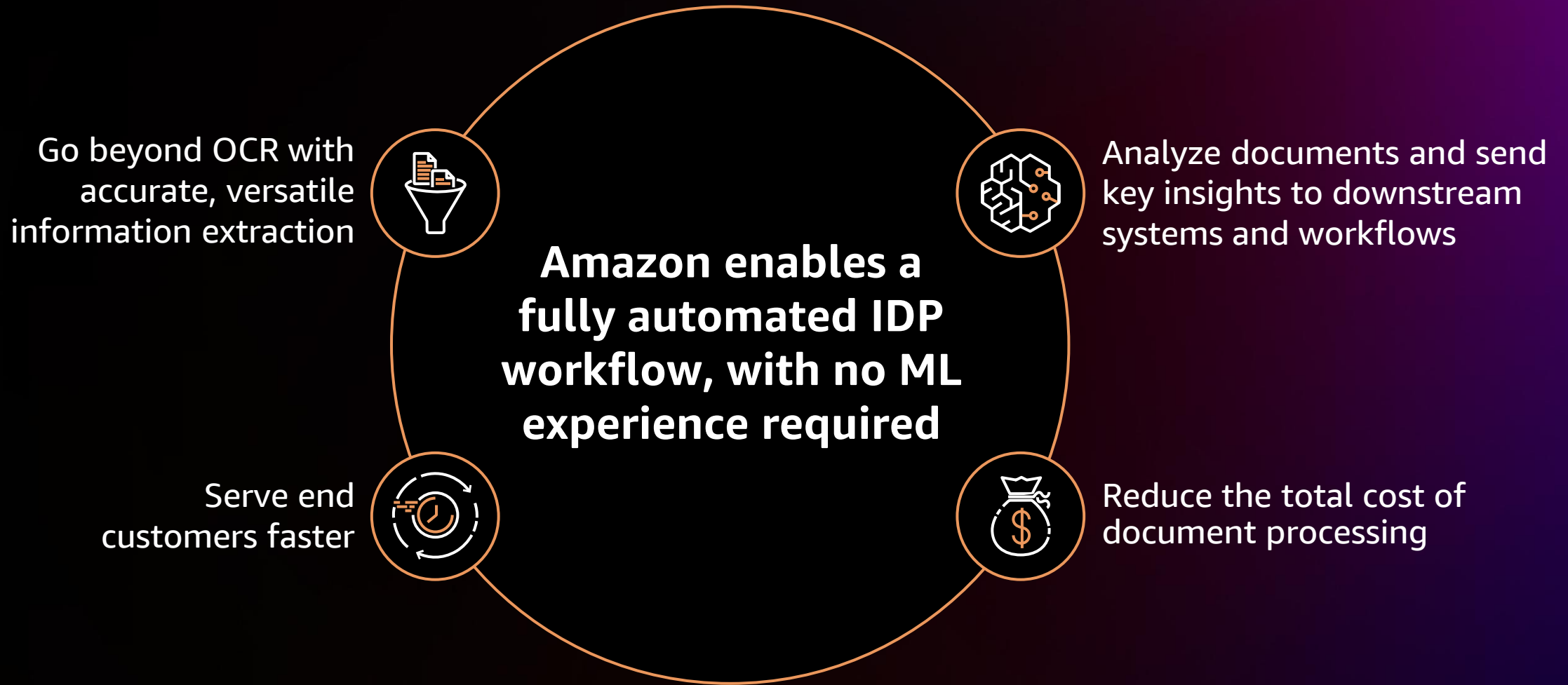
Higher costs/risks

Staff unavailable for higher value-add activities vs. "keeping lights on"

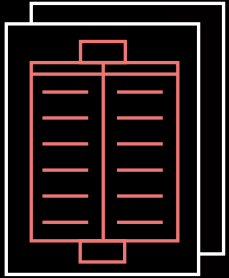
Lower quality decisions caused by error-prone systems

Technical debt that leads to security risks, compliance issues, and lack of business continuity

Benefits of Amazon Textract



Amazon Textract features

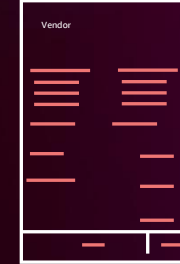


Text (OCR)

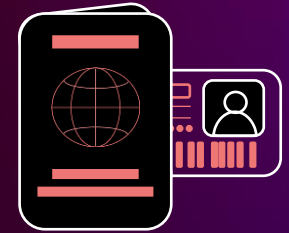


Forms

Invoices & receipts



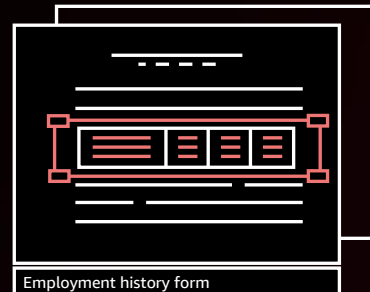
Identity documents



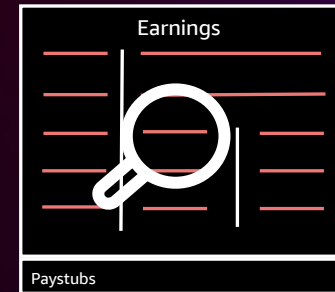
Specialized documents



Handwriting



Tables



Queries

Data extraction challenges



Data variations

(DOB vs. date of birth
vs. birthdate)

Post-processing

Expensive

Time-consuming



Data structure

(Table, form,
implied fields)

Post-processing

Custom models

Manual



Nested data

(page sections,
duplicate fields)

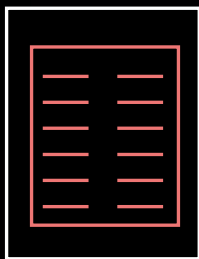
Development and
management overhead

Templates are brittle

Amazon Textract Queries

Easily specify and extract valuable pieces of information from documents

DOCUMENT



QUESTIONS

Sample questions:

1. What is the borrower's tax identifier?
2. What are YTD earnings in this paystub?
- ... and more



OUTPUT

1. What is the borrower's tax identifier? [borrower tax id]
2. What are YTD earnings in this paystub? [YTD earnings]
- ...

**No custom model
training or synonym
lists required**



Combines visual, spatial, and textual cues to provide higher accuracy



Simple Q&A response reduces post-processing



Extract only the specific fields that matter



Easy integration with existing Amazon Textract Analyze Document API

Amazon Textract Queries on Fannie Mae Form 1003

NESTED SECTIONS

Uniform residential loan application (Form 1003)

Borrower				III. BORROWER INFORMATION				Co-Borrower							
Borrower's Name (include Jr. or Sr. if applicable)				Co-Borrower's Name (include Jr. or Sr. if applicable)											
Social Security Number		Home Phone (incl. area code)		DOB (mm/dd/yyyy)		Yrs. School		Social Security Number		Home Phone (incl. area code)		DOB (mm/dd/yyyy)		Yrs. School	
				1/1/1900		18						2/2/1902		18	
☒ Married ☐ Separated		☐ Unmarried (include single, divorced, widowed)		Dependents (not listed by Co-Borrower)				☒ Married ☐ Separated		☐ Unmarried (include single, divorced, widowed)		Dependents (not listed by Borrower)			
				no.		ages						no.		ages	

For example

- What is the borrower's date of birth? **01/01/1900**
- What is the co-borrower's date of birth? **02/02/1902**

Amazon Textract Queries on bank statements

IMPLIED FIELDS

Bank statement

RETURN SERVICE REQUESTED

John Doe
123 Main Street
Baltimore, MD 21224

Managing Your Accounts

	Primary Branch	Canton
	Phone Number	443-573-4800
	Online Banking	HowardBank.com
	Telephone Banking	1-877-527-2703
	Mailing Address	3301 Boston Street Baltimore, MD 21224

For example:

- What is the customer name? **John Doe**
- What is the customer address? **123 Main Street, Baltimore, MD 21224**

Amazon Textract Queries on paystubs

TABULAR DATA

Pay stubs

CO. FILE DEPT. CLOCK NUMBER ABC 126543 123456 12345 00000000 1					Earnings Statement	
ANY COMPANY CORP. 475 ANY AVENUE ANYTOWN, USA 10101					Period ending:	7/18/2008
					Pay date:	7/25/2008
Social Security Number: 987-65-4321 Taxable Marital Status: Married Exemptions/Allowances: Federal: 3, \$25 Additional Tax State: 2 Local: 2					JOHN STILES 101 MAIN STREET ANYTOWN, USA 12345	
Earnings	rate	hours	this period	year to date	Other Benefits and Information	
Regular	10.00	32.00	320.00	18,840.00	Group Term Life	0.51 27.00
Overtime	15.00	1.00	15.00	760.00	Loan Amt Paid	840.00
Holiday	10.00	8.00	80.00	4,160.00	Vac Hrs	40.00
Tuition			37.43*	1,948.80	Sick Hrs	16.00
					Title	Operator
Gross Pay			\$ 452.43	23,526.80	Important Notes	
Deductions	Statutory				EFFECTIVE THIS PAY PERIOD YOUR REGULAR HOURLY RATE HAS BEEN CHANGED FROM \$8.00 TO \$10.00 PER HOUR.	
	Federal Income Tax		- 40.80	2,111.20	WE WILL BE STARTING OUR UNITED WAY FUND DRIVE SOON AND LOOK FORWARD TO YOUR PARTICIPATION.	
	Social Security Tax		- 28.05	1,458.60		
	Medicare Tax		- 6.56	341.12		
	NY State Income Tax		- 8.43	438.36		
	NYC Income Tax		- 5.94	308.88		
	NY SUI/SDI Tax		- 0.60	31.20		
	Other					
	Bond		- 5.00	100.00		
	401(k)		- 28.85*	1,500.20		
	Stock Plan		-15.00	150.00		
	Life Insurance		- 5.00	50.00		
	Loan		- 30.00	150.00		
	Adjustment					
	Life Insurance		+ 13.50			
Net Pay			\$ 291.90			
* Excluded from federal taxable wages						
Your federal wages this period are \$386.15						

For example

- What is the gross pay this period? **\$452.43**
- What is the gross pay YTD? **\$23,526.80**
- What is the Medicare tax YTD? **\$341.12**

Packaged use-case solution to accelerate workflows : Analyze Lending

AMAZON TEXTTRACT DOCUMENT PROCESSING SOLUTION FOR MORTGAGES

The image shows a sample of a Uniform Residential Loan Application form. It includes sections for Earnings Statement, Closing Disclosure, and the main application form. The application form has various fields for borrower information, loan details, and property information. A large arrow points from the form towards the 'Analyze Lending API' box.

Analyze Lending API

Classify & split mortgage documents

Extract data

Validate mortgage package



Automated end-to-end processing of mortgage packages



Pre-trained on diverse sets of mortgage document types



Faster time-to-value



Learn how Black Knight is using AI to accelerate mortgage workflows



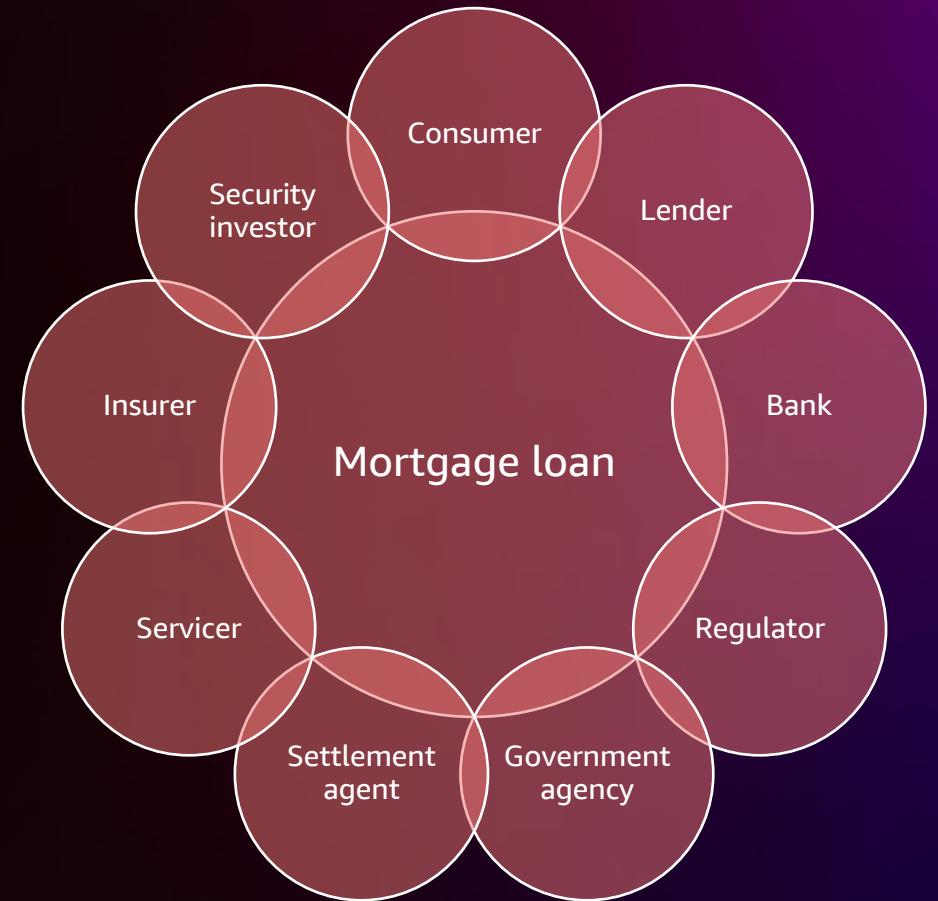
Black Knight (NYSE:BKI)

- **A premier provider of technology solutions for mortgage and home equity lending and servicing**
- **Solutions include:**
 - Loan origination – Empower
 - Loan servicing – MSP
 - Mortgage capital markets – Optimal Blue
 - Data and analytics marketplace for mortgage, real estate, and capital markets
- **AIVA is our AI/ML-based document service**

The real estate finance industry

- **The residential mortgage lending market in 2021**
 - 4,316 financial institutions¹
 - Loans originated – \$4.5 trillion, 13.5 million loans²
 - Loans serviced – \$12.5 trillion, 52.4 million loans²
- **The mortgage value stream spans up to 47 years**
- **The industry is complex – many stakeholders**

Mortgage loan stakeholders



Paper still dominates

- **Residential mortgages require hundreds of documents**
 - 800+ document types and 600+ unique document/label pairs (and counting)
 - Many documents are highly variable and few are standardized
- **Most are scanned from paper documents**
 - 24% of borrowers get important initial documents on paper
 - Despite 23-year-old UETA law, only about 4% of loans are closed digitally
- **It's expensive – total cost to originate a loan is \$10,937¹**

AIVA document services

The mission: classify 800+ document types, extract 600+ name/value pairs

The solution

- AWS infrastructure + Amazon Textract + Black Knight proprietary AI/ML models, software, and data
- Humans-in-the-loop to manage quality, extract ground truth
- Key strategies
 - Classifier – open-source OCR + Black Knight-trained, transformer-based NLP machine learning model
 - Extraction – focused on Amazon Textract knowledge graph and proprietary models
 - Business rules – accelerating processes based on documents received and data extracted

Example: Solving for bank statements

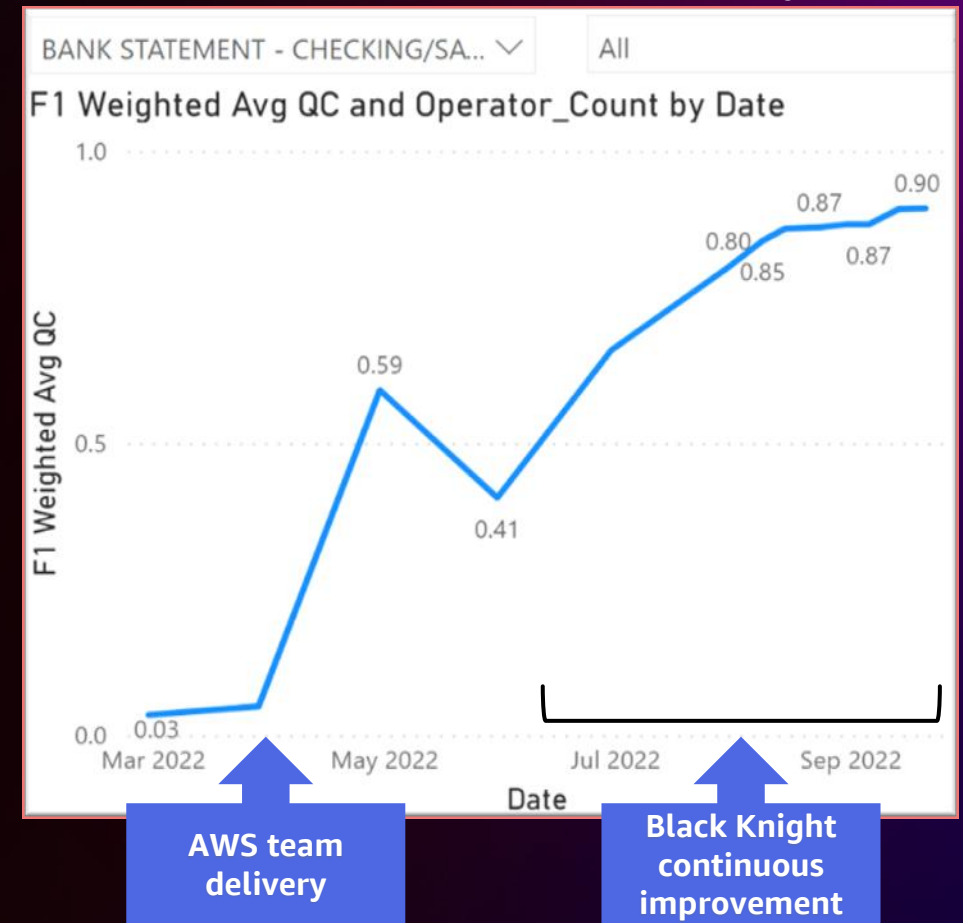
- **Challenge**

- Many varieties – 9,624 financial institutions in US
- Every consumer is different, every month
- Data spans pages, often hundreds of rows of bank transactions

- **Solved by Black Knight data science team and Amazon Textract experts**

- **Output accuracy now 90%+ (F1 measure)**

Weighted average F_1
(a measure of model accuracy)



AIVA value

- **Document services at scale**
 - Processing 14 million+ pages/month
 - Extracting 7 million+ datapoints/month
- **Clients report significant process cost and time savings**
 - 5–10 seconds per page
 - Seconds to minutes per datapoint
 - Data-driven process can save hours per loan
- **Adoption is accelerating**

AWS Partners can help accelerate IDP projects

				Suggest build in-house	Suggest partner
					
				X	
					X
					
					X
				X	
Document processing is a top corporate competency					
Difficulty hiring enough experienced AWS developers with knowledge of AI/ML					
Using a partner for digital transformation projects					
Already use serverless architecture to manage document workflows					

This is not a complete list. To view all AWS Partners for this category, visit [AWS Partner Solutions Finder](#). This list of partners is current as of Oct. 17, 2022.

Important links

Service links

- [Amazon Textract](#)
- [Amazon Comprehend](#)
- [Amazon Augmented AI \(Amazon A2I\)](#)

Blog posts

- [Build an end-to-end intelligent document processing solution using AWS](#)
- [Automating Customer Onboarding with Lumiq's Drishti Document AI and Amazon Textract](#)
- [Broadridge Advances the Proxy Voting Data Process Using Artificial Intelligence on AWS](#)

Git source code link

- [Document understanding solution](#)
- [Amazon Textract samples](#)
- [Amazon Comprehend samples](#)
- [Amazon A2I samples](#)

Account team can provide

- Amazon Textract and Amazon Comprehend security papers (can be shared under NDA)
- Immersion day training
- Partner referrals

Thank you!



Please complete the session survey in the **mobile app**

